

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-13832

(43) 公開日 平成10年(1998) 1月16日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 N 7/30			H 0 4 N 7/133	Z
G 0 6 T 7/20			G 0 6 F 15/70	4 1 0

審査請求 未請求 請求項の数16 O L (全 13 頁)

(21) 出願番号 特願平8-164430

(22) 出願日 平成8年(1996) 6月25日

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 大和 淳司

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72) 発明者 村瀬 洋

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74) 代理人 弁理士 秋田 収喜

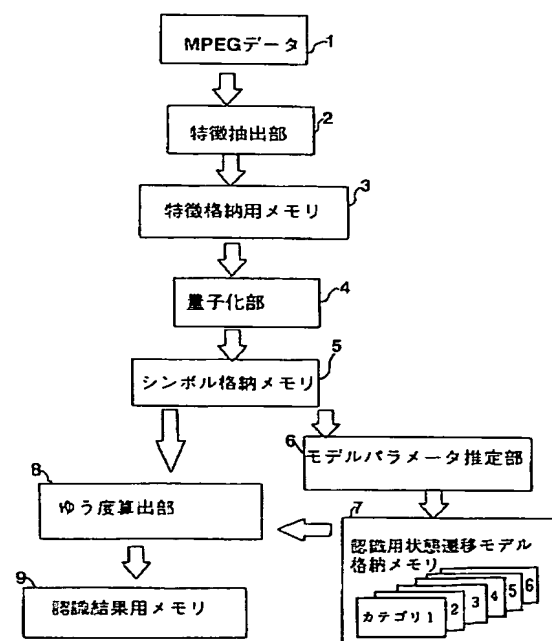
(54) 【発明の名称】 動画像認識方法および動画像認識検索方法

(57) 【要約】

【課題】 標準符号化方式等により圧縮された動画像データの使い、データ処理時間を短縮した動画像認識検索方法を提供する。

【解決手段】 一連の動画像を表示する各画面の画像データをM×Nのブロックに区切り、各ブロックのDCT係数を抽出するステップと、前記DCT係数の少なくとも1つを各画面の特徴ベクトルとして抽出するステップと、検索キーとなる特定の動画像パターンを表示する各画面の特徴ベクトルで構成される時系列の特徴ベクトル列により、確率的な状態遷移モデルを学習するステップと、検索対象である一連の動画像を表示する各画面の画像データから抽出された特徴ベクトルで構成される時系列の特徴ベクトル列の中で、前記学習により得られた状態遷移モデルに対する尤度が高い時間領域を検索結果として出力するステップとを具備する。

図 1



【特許請求の範囲】

【請求項 1】 一連の動画像の動画像パターンを認識する動画像認識方法において、一連の動画像を表示する各画面の画像データを $M \times N$ のブロックに区切り、各ブロックの DCT 係数を抽出するステップと、前記 DCT 係数の少なくとも 1 つを各画面の特徴ベクトルとして抽出するステップと、特定の動画像パターンを表示する各画面の特徴ベクトルで構成される時系列の特徴ベクトル列により、確率的な状態遷移モデルを、認識キーとなる複数の特定の動画像パターン毎に学習するステップと、認識対象である一連の動画像を表示する各画面の画像データから抽出された特徴ベクトルで構成される時系列の特徴ベクトル列の、前記学習により得られた複数の状態遷移モデルに対する尤度が最大となる状態遷移モデルの動画像パターンを認識結果として出力するステップとを具備することを特徴とする動画像認識方法。

【請求項 2】 前記認識対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部を使用することを特徴とする請求項 1 に記載された動画像認識方法。

【請求項 3】 前記各画面の特徴ベクトルとして、DCT 係数とともに動きベクトルを使用することを特徴とする請求項 1 に記載された動画像認識方法。

【請求項 4】 前記認識対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部、および、動き補償ベクトルを使用することを特徴とする請求項 3 に記載された動画像認識方法。

【請求項 5】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、3 ないし 21 個の低周波成分の DCT 係数を特徴ベクトルして使用することを特徴とする請求項 2 または請求項 4 に記載された画像認識方法。

【請求項 6】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、水平方向の 1 番目のライン上の DCT 係数を特徴ベクトルして使用することを特徴とする請求項 2 または請求項 4 に記載された画像認識方法。

【請求項 7】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、垂直方法の 1 番目のライン上の DCT 係数を特徴ベクトルして使用することを特徴とする請求項 2 または請求項 4 に記載された画像認識方法。

【請求項 8】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、直流成分を含む対角線上の DCT 係数を特徴ベクトルして使用

することを特徴とする請求項 2 または請求項 4 に記載された画像認識方法。

【請求項 9】 一連の動画像の中から、特定の動画像パターンを含む時間領域を抽出する動画像認識検索方法において、一連の動画像を表示する各画面の画像データを $M \times N$ のブロックに区切り、各ブロックの DCT 係数を抽出するステップと、前記 DCT 係数の少なくとも 1 つを各画面の特徴ベクトルとして抽出するステップと、検索キーとなる特定の動画像パターンを表示する各画面の特徴ベクトルで構成される時系列の特徴ベクトル列により、確率的な状態遷移モデルを学習するステップと、検索対象である一連の動画像を表示する各画面の画像データから抽出された特徴ベクトルで構成される時系列の特徴ベクトル列の中で、前記学習により得られた状態遷移モデルに対する尤度が高い時間領域を検索結果として出力するステップとを具備することを特徴とする動画像認識検索方法。

【請求項 10】 前記検索対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部を使用することを特徴とする請求項 9 に記載された動画像認識検索方法。

【請求項 11】 前記各画面の特徴ベクトルとして、DCT 係数とともに動きベクトルを使用することを特徴とする請求項 9 に記載された動画像認識検索方法。

【請求項 12】 前記検索対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部、および、動き補償ベクトルを使用することを特徴とする請求項 11 に記載された動画像認識検索方法。

【請求項 13】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、3 ないし 21 個の低周波成分の DCT 係数を特徴ベクトルして使用することを特徴とする請求項 10 または請求項 12 に記載された画像認識検索方法。

【請求項 14】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、水平方向の 1 番目のライン上の DCT 係数を特徴ベクトルして使用することを特徴とする請求項 10 または請求項 12 に記載された画像認識検索方法。

【請求項 15】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、垂直方法の 1 番目のライン上の DCT 係数を特徴ベクトルして使用することを特徴とする請求項 10 または請求項 12 に記載された画像認識検索方法。

【請求項 16】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、直流

成分を含む対角線上の DCT 係数を特徴ベクトルとして使用することを特徴とする請求項 10 または請求項 12 に記載された画像認識検索方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、動画画像認識方法および動画画像認識検索方法に係わり、特に、一連の動画画像を表示する各画面の画像データの中から特定の動画画像パターンを認識・検索を行う動画画像認識方法および動画画像認識検索方法に関する。

【0002】

【従来の技術】動画画像を対象としたパターン認識技術は、近年多くの研究が行われており、その一つとして、下記公報 (イ) に記載されている手法が公知である。

【0003】(イ) 特開平 5-46583 号公報
前記公報 (イ) (特開平 5-46583 号公報) には、動画画像を表示する各画面の画像データから抽出した動物体のメッシュ特徴をベクトル量子化によりシンボル化して、動画画像系列をシンボル系列に変換し、当該シンボル系列を学習・認識することにより、人間等の動物体の各動作を認識する手法が記載されている。

【0004】また、マルチメディアの中核技術を構成する、動画画像データの蓄積あるいは伝送の際の情報圧縮技術として、MEPG (Moving Picture Experts Group; メディア統合系動画画像圧縮の国際標準)、MEPG2 といった国際標準符号化方式が普及しつつある。

【0005】

【発明が解決しようとする課題】前記公報 (イ) (特開平 5-46583 号公報) に記載されている手法のように、従来一連の動画画像の中から、特定の動画画像パターンをその動画画像パターン自体を検索キーとして検索する場合は、大容量の画像データおよび特徴量データを取り扱う必要があり、データ処理の処理時間が増大するという問題点があった。

【0006】また、MEPG、MEPG2 等の標準符号化方式が普及しつつあり、一連の動画画像の中から、特定の動画画像パターンをその動画画像パターン自体を検索キーとして検索する場合に、この標準符号化方式により圧縮された動画画像データを使用することにより、データ処理の処理時間を短縮することが期待される。

【0007】しかしながら、標準符号化方式により圧縮された動画画像データを対象として、一連の動画画像の中から特定の動画画像パターンを検索する最適な手法について、従来何ら検討されていなかった。

【0008】本発明は、前記問題点を解決するためになされたものであり、本発明の目的は、動画画像認識方法において、標準符号化方式等により圧縮された動画画像データを使用し、データ処理時間を短縮することが可能となる技術を提供することにある。

【0009】本発明の他の目的は、動画画像認識検索方法において、標準符号化方式等により圧縮された動画画像データを使用し、データ処理時間を短縮することが可能となる技術を提供することにある。

【0010】本発明の前記目的並びにその他の目的及び新規な特徴は、本明細書の記載及び添付図面によって明らかにする。

【0011】

【課題を解決するための手段】本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、下記の通りである。

【0012】(1) 一連の動画画像の動画画像パターンを認識する動画画像認識方法において、一連の動画画像を表示する各画面の画像データを $M \times N$ のブロックに区切り、各ブロックの DCT 係数を抽出するステップと、前記 DCT 係数の少なくとも 1 つを各画面の特徴ベクトルとして抽出するステップと、特定の動画画像パターンを表示する各画面の特徴ベクトルで構成される時系列の特徴ベクトル列により、確率的な状態遷移モデルを、認識キーとなる複数の特定の動画画像パターン毎に学習するステップと、認識対象である一連の動画画像を表示する各画面の画像データから抽出された特徴ベクトルで構成される時系列の特徴ベクトル列の、前記学習により得られた複数の状態遷移モデルに対する尤度が最大となる状態遷移モデルの動画画像パターンを認識結果として出力するステップとを具備することを特徴とする。

【0013】(2) 前記 (1) の手段において、前記認識対象である一連の動画画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部を使用することを特徴とする。

【0014】(3) 前記 (1) の手段において、前記各画面の特徴ベクトルとして、DCT 係数とともに動きベクトルを使用することを特徴とする。

【0015】(4) 前記 (3) の手段において、前記認識対象である一連の動画画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部、および、動き補償ベクトルを使用することを特徴とする。

【0016】(5) 一連の動画画像の中から、特定の動画画像パターンを含む時間領域を抽出する動画画像認識検索方法において、一連の動画画像を表示する各画面の画像データを $M \times N$ のブロックに区切り、各ブロックの DCT 係数を抽出するステップと、前記 DCT 係数の少なくとも 1 つを各画面の特徴ベクトルとして抽出するステップと、検索キーとなる特定の動画画像パターンを表示する各画面の特徴ベクトルで構成される時系列の特徴ベクトル列により、確率的な状態遷移モデルを学習するステップ

と、検索対象である一連の動画像を表示する各画面の画像データから抽出された特徴ベクトルで構成される時系列の特徴ベクトル列の中で、前記学習により得られた状態遷移モデルに対する尤度が高い時間領域を検索結果として出力するステップとを具備することを特徴とする。

【0017】(6)前記(5)の手段において、前記検索対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の一部を使用することを特徴とする。

【0018】(7)前記(5)の手段において、前記各画面の特徴ベクトルとして、DCT係数とともに動きベクトルを使用することを特徴とする。

【0019】(8)前記(7)の手段において、前記検索対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の一部、および、動き補償ベクトルを使用することを特徴とする。

【0020】前記各手段によれば、特徴量としてDCT係数、あるいはDCT係数および動き補償ベクトルを使用し、MEPG、MEPG2等の標準符号化方式で圧縮された少量の動画像データから、特定の動画像パターンを直接認識・検索するようにしたので、データ処理の処理時間を少なくすることが可能となる。

【0021】

【発明の実施の形態】以下、本発明の発明の実施の形態を図面を参照して詳細に説明する。

【0022】なお、発明の実施の形態を説明するための全図において、同一機能を有するものは同一符号を付け、その繰り返しの説明は省略する。

【0023】図1は、本発明の一発明の実施の形態である動画像認識方法および動画像認識検索方法が適用される動画像認識検索装置の概略構成を示す機能ブロック図である。

【0024】図1において、1はMEPGデータ、2は特徴抽出部、3は特徴格納用メモリ、4は量子化部、5はシンボル格納メモリ、6はモデルパラメータ推定部、7は認識用状態遷移モデル格納メモリ、8は尤度算出部、9は認識結果用メモリである。

【0025】ここで、前記認識用状態遷移モデル格納メモリ7および認識結果用メモリ9としては、例えば、外部記憶装置を使用し、また、前記MEPGデータ1は、例えば、外部記憶装置に格納されている。

【0026】本発明の実施の形態の基本的動作には、学習と認識の3つの段階があり、学習時には、学習用のデータから認識用状態遷移モデルのパラメータ推定を行い認識カテゴリ(図1に示すカテゴリ1～カテゴリ6)毎に認識用状態遷移モデル格納メモリ7に格納する。

【0027】また、認識時には、学習によって認識用状態遷移モデル格納メモリ7に格納された、各カテゴリに対応するモデルの尤度を算出し、最大の尤度を持つモデルに対応するカテゴリを認識結果とする最尤推定を行う。

【0028】本発明の実施の形態の動画像認識方法および動画像認識検索方法において、量子化までの処理は学習時、認識時とも同一である。

【0029】以下、図1に沿って、本発明の実施の形態の動画像認識方法および動画像認識検索方法について説明する。

【0030】まず、検索対象のMEPGデータ1から、特徴抽出部2により、特徴ベクトルとして、DCT係数を抽出する。

【0031】ここで、MEPGデータ1について、簡単に説明する。

【0032】MEPG標準化符号化方式では、フレーム内では8×8画素のブロック毎のDCT(離散コサイン変換; Discrete Cosine Transform)係数と量子化により、また、フレーム間では動き補償ベクトル情報を使用して、データを圧縮している。

【0033】また、通常のMEPGデータ1の各フレームは、Iピクチャ、Pピクチャ、Bピクチャの3種類のいずれかのタイプの符号化データで構成される。

【0034】なお、Iピクチャはフレーム内符号化、Pピクチャは順方向フレーム間予測符号化、Bピクチャは双方向フレーム間予測符号化を意味する。

【0035】通常のシーケンスでは、1つのGOP(Group of Picture)は、Iピクチャで始まり、画像の動きの激しさや要求画質等に応じて、適当な間隔でPピクチャあるいはBピクチャを配置する。

【0036】本発明の実施の形態では、DCT係数を利用するために、全てのフレームをIピクチャである画像データに変換して使用する。

【0037】なお、I、P、Bピクチャから構成されるMEPGデータ1からIピクチャへのへの変換は、例えば、下記文献(ロ)に記載されているように、符号化データを直接操作することにより可能である。

【0038】(ロ) Shin-Fu Chang and David G. Messerschmitt: "A New Approach to Decoding and Compositing Motion-Compensated DCT-Based Images", Proceedings of ICASSP '93(1993). 図2は、MEPGデータ1およびMEPGデータ1のDCT係数の概略構成を示す図である。

【0039】図2に示すように、MEPGデータ1では、1フレームの画像データを、1ブロックが8×8画素で構成されるM×Nブロックに分割し、そのブロック単位にDCT演算を行い、それにより、図2の最下段のブロック内の数字1～64に示すDCT係数が得られる。

【0040】本発明の実施の形態では、この 8×8 画素のブロックのDCT係数の内、低周波成分のDCT係数（図3に示すE1の領域のDCT係数）を適当な数を取り出し、これを全ブロックに対して行い、取り出されたDCT係数全体を並べた数値列を、そのフレームの特徴ベクトル（f）とする。

【0041】仮に、 32 画素 $\times 32$ 画素の画像を使用し、各ブロックから i 個のDCT係数を取り出すとすると、全部で 16 個のブロックがあるので、この場合の特徴ベクトルの次元は $16i$ となる。

【0042】MPEGデータ1の1フレームの画像データから1つの特徴ベクトル（f）が得られるので、一連の動画像を表示する連続したフレーム（画面）の画像データから特徴ベクトル列（F）が得られ、この特徴ベクトル列（F）は、特徴格納用メモリ3に記録される。

【0043】なお、特徴ベクトル（f）として使用するDCT係数は、低周波成分の適当な数のDCT係数以外に、水平方向の1番目のライン上のDCT係数（図3に示すE2の領域のDCT係数）、垂直方向の1番目のライン上のDCT係数（図3に示すE3の領域のDCT係数）、あるいは、直流成分を含む対角線上のDCT係数（図3に示すE4の領域のDCT係数）を使用するようにしてもよい。

【0044】水平方向の1番目のライン上のDCT係数（図3に示すE2の領域のDCT係数）を特徴ベクトルとして使用することにより、動画像の特定パターンが主に水平方向の動きが支配的である場合に、少ないDCT係数で精度よく動画像の特徴を抽出することが可能である。

【0045】また、垂直方向の1番目のライン上のDCT係数（図3に示すE3の領域のDCT係数）を特徴ベクトルとして使用することにより、動画像の特定パターンが主に垂直方向の動きが支配的である場合に、少ないDCT係数で精度よく動画像の特徴を抽出することが可能である。

$$C = c_1, c_2, \dots, c_N$$

【0054】

【数2】 $O_i = v_k \dots \dots (2)$

$k = \arg \min_j d(f, c_j)$

但し、 $d(x, y)$ は x, y の距離

ここまでの処理によって、特徴ベクトル列（F）がシンボル列（O）に変換され、このシンボル列（O）を、状態遷移モデルにより、学習、認識を行う。

【0055】なお、ここまでの動作については、認識時、学習時ともに同一である。

【0056】この状態遷移モデルとしては、前記文献（二）、あるいは、下記文献（ホ）に記載されている隠れマルコフ（以下、HMMと称す。）モデルを使用する。

【0057】（ホ）中川聖一；“確率モデルによる音声

【0046】また、直流成分を含む対角線上のDCT係数（図3に示すE4の領域のDCT係数）を特徴ベクトルとして使用することにより、動画像の特定パターンが水平方向および垂直方向の動きの両方を含んでいる場合に、少ないDCT係数で精度よく動画像の特徴を抽出することが可能である。

【0047】さらに、特徴ベクトル（f）としては、DCT係数と動き補償ベクトルとを併用することも可能であり、これにより、より詳細に動画像の特徴を抽出することが可能となる。

【0048】この特徴ベクトル列（F）は、量子化部4でベクトル量子化によって、シンボル列（O）へ変換され、シンボル格納メモリ5に記録される。

【0049】即ち、各特徴ベクトルはあらかじめ用意された量子化のための代表点の一覧に基づき、それらの内で最も距離の近い代表点ベクトルに対応するシンボルに変換される。

【0050】この代表点群をコードブックと呼び、このコードブックは、各種の動作画像から抽出した特徴ベクトルの一部を利用して、下記文献（ハ）に記載されているLBGアルゴリズムで作成した。

【0051】（ハ）Y.Linde, A.Buzo, R.M.Gray; "An Algorithm for Vector Quantizer design", IEEE Trans. Comm. vol.COM-28 (1980). なお、このコードブックの作成には、下記文献（二）に記載されているk-means (k-平均) アルゴリズムで作成してもよい。

【0052】（二）X.D.Huang, Y.Ariki, M.A.Jack; "Hidden Markov Model for Speech Recognition", Edinburgh Univ. Press (1990). 今、コードブックを下記

（1）式のように表現するとすると、特徴ベクトル（f）は、下記式（2）に示すシンボル（ O_i ）に変換される。

【0053】

【数1】

$$\dots \dots (1)$$

認識”，電子情報通信学会（1990）

学習時には、前記HMMモデルのパラメータの推定を行い、また、認識時には、認識するカテゴリ数だけ用意され、認識用状態遷移モデル格納メモリ7に格納されたHMMモデルの各々から、認識対象の特徴ベクトル列（F）が生成される確率を尤度算出部8によって算出する。

【0058】以下、HMMモデルについて、簡単に説明する。

【0059】HMMモデルは、確率的な状態遷移モデルであり、時系列現象の発生源のモデル化と見ることがができる。

【0060】図4は、HMMモデルの概念を示す概念図である。

【0061】図4に示すように、HMMモデルには、複数の状態 ($q_1 \sim q_5$) が存在し、各状態 ($q_1 \sim q_5$) から他の状態へ遷移する確率 (a_{ij}) が与えられている。

【0062】時刻が進につれて状態遷移が確率的に発生し、さらに、各状態から確率的にシンボル ($O_1 \sim O_t$) が出力される。

【0063】観測可能なのはこの出力シンボル列 ($O = O_1, O_2, \dots, O_t$) であり、状態を直接観測することができない。

【0064】これが“隠れ”マルコフモデルの由来である。

【0065】動作認識への適用においては、動作中における各姿勢が状態に当たり、従って、状態数は、認識対象動作の長さや複雑さに応じて適当な数を選択する必要がある。

【0066】また、動作認識への適用においては、状態遷移確率が姿勢変化の時系列パターン自身とその伸縮などの変化を、シンボル出力確率が、各姿勢の揺らぎや姿勢の観測結果の揺らぎを記述する部分に当たると解釈できる。

【0067】HMMモデルは、以下のパラメータで記述される。

【0068】

【数3】 $S = \{s_i\}$: 状態の集合。 s_i は t 番目の状態 (観測できない)

$O = O_1, O_2, \dots, O_T$: 観測されたシンボル系列 (長さ T)

$A = \{a_{ij} \mid a_{ij} = \Pr(s_{t+1} = j \mid s_t = i)\}$: 状態遷移確率

a_{ij} は状態 (s_i) から状態 (s_j) へ遷移する確率

$B = \{b_j(O_t) \mid b_j(O_t) = \Pr(O_t \mid s_t = j)\}$: シンボル出力確率

$b_j(k)$ は状態 (s_j) においてシンボル (v_k) を出力する確率

モデル $\lambda = (\pi_i, a_{ij}, b_i(v))$ から、より良いモデル $\bar{\lambda} = (\bar{\pi}_i, \bar{a}_{ij}, \bar{b}_i(v))$ を、次のようにして求める。

【0078】

【数5】

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}; \quad \dots \dots (3)$$

$$\bar{b}_i(k) = \frac{\sum_{t \in \{t \mid O_t = v_k\}} \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)}; \quad \dots \dots (4)$$

【0080】

【数7】

$$\bar{\pi}_i = \gamma_1(i). \quad \dots \dots (5)$$

$\pi = \{\pi_i \mid \pi_i = \Pr(s_1 = i)\}$: 初期状態確率

次に、HMMモデルを使用した時系列パターン (シンボル列 (O)) の学習と認識の手順について説明する。

【0069】《学習時の手順》モデルパラメータ推定部6は、各カテゴリ毎に複数与えられた学習用データから得られたシンボル列 (O) に対して、そのシンボル列

(O) を発生するような状態遷移モデルのパラメータを推定し、認識用状態遷移モデル格納メモリ7に蓄える。

【0070】HMMモデルによる認識系は、各カテゴリ毎に1つのHMMモデルから構成される。

【0071】いま、認識対象の各カテゴリ毎のHMMモデルを $\lambda_i (= \{A_i, B_i, \pi_i\})$ とすると、この λ_i の学習を各カテゴリ毎の学習パターンを用いて行う。

【0072】ここで、学習とは、学習パターンを発生しやすようなHMMモデルのパラメータ、即ち、状態遷移確率 A_i 、シンボル出力確率 B_i と初期状態確率 π_i を推定することに他ならない。

【0073】学習パターンからHMMモデルのパラメータを推定するには、前記文献(二)、あるいは、文献(ホ)に記載されているBauner-Welchアルゴリズムを使用する。

【0074】具体的には、ある初期値から順に、より尤度の高いHMMモデルのパラメータを求めることを、尤度の値、変化などから充分収束したと見なせるまで繰り返す手続き、即ち、あるHMMモデルのパラメータをもとに、それよりもより尤度の高いモデルパラメータを求めることを繰り返していく手続きである。

【0075】繰り返し毎に、前記文献(二)に記載されているforwardアルゴリズムによって尤度の値を確認することで収束の確認が可能である。

【0076】数式で表現すると、

【0077】

【数4】

【0079】

【数6】

【0081】但し、ここで、

【0082】

【数8】

11

12

$$\begin{aligned}
 \gamma_t(i) &\equiv \frac{P(s_t = i | O_1, O_2, \dots, O_T, \lambda)}{P(s_t = i, O_1, O_2, \dots, O_T | \lambda)} \\
 &= \frac{P(O_1, O_2, \dots, O_T | \lambda)}{P(O_1, O_2, \dots, O_t, s_t = i | \lambda) P(s_t = i, O_{t+1}, O_{t+2}, \dots, O_T | \lambda)} \\
 &= \frac{\alpha_t(i) \beta_t(i)}{P(O | \lambda)} \dots \dots \dots (6)
 \end{aligned}$$

【0083】

10 【数9】

$$\begin{aligned}
 \xi_t(i, j) &\equiv \frac{P(s_t = i, s_{t+1} = j | O_1, O_2, \dots, O_T, \lambda)}{P(s_t = i, s_{t+1} = j, O_1, O_2, \dots, O_T | \lambda)} \\
 &= \frac{P(O_1, O_2, \dots, O_T | \lambda)}{P(O_1, O_2, \dots, O_t, s_t = i | \lambda) a_{ij} b_j(O_{t+1}) P(O_{t+2}, O_{t+3}, \dots, O_T | \lambda)} \\
 &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)} \dots \dots \dots (7)
 \end{aligned}$$

【0084】前記各式の意味するところは、(3)式は、HMMモデルλのもとでの a_{ij} の再評価であり、

(4)式は、HMMモデルλのもとでの $b_i(k)$ の再評価である。

【0085】前記した手続きによって、学習データに対応する認識用状態遷移モデルのパラメータを求めることができる。

【0086】こうして求めた各カテゴリ毎のモデルを認識の際に使用する。

【0087】《認識時の手順》認識の手順は、各HMMモデルの尤度計算と最大値の選択で行われる。

【0088】認識対象のパターンに対して、 λ_i が、認識対象パターンであるシンボル列($O = O_1, O_2, \dots, O_t$)を出力する確率(尤度) $Pr(O | \lambda_i)$ を計算する。

$$\alpha_t(i) \equiv Pr(O_1, O_2, \dots, O_t, s_t = i | \lambda) \dots \dots \dots (9)$$

【0094】で定義される値で、HMMモデルλがシンボル系列($O = O_1, O_2, \dots, O_t$)を発生して、時間tにおいて状態($S_t = i$)である確率である。

$$\alpha_t(j) = \left\{ \sum_i \alpha_{t-1}(i) a_{ij} \right\} b_j(O_t) \dots \dots \dots (10)$$

$$\alpha_1 = \pi_i b_i(O_1) \dots \dots \dots (11)$$

【0097】の漸化式で求められる。

【0098】こうして求められた尤度が最大となるモデル、即ち、式(1)から式(11)で求めた $Pr(O | \lambda_i)$ から、尤度最大の λ_i に対するカテゴリ(G_k)

20 【0089】尤度の計算は、前記文献(二)に記載されているforwardアルゴリズムによって再帰的に、以下のようにして求めることができる。

【0090】即ち、あるモデル $\lambda = \{A, B, \pi\}$ がシンボル系列($O = O_1, O_2, \dots, O_t$)を出力する確率 $Pr(O | \lambda_i)$ は、

【0091】

【数10】

$$Pr(O | \lambda) = \sum_{i \in S_F} \alpha_T(i) \dots \dots \dots (8)$$

30 【0092】ただし、ここで、 S_F は最終状態の集合、 $\alpha_T(i)$ は、

【0093】

【数11】

【0095】これは、

【0096】

【数12】

($k = \arg \max_i Pr(O | \lambda_i)$)が認識結果として選択され、認識結果用メモリ6に蓄えられる。

【0099】また、検索時には、検索対象となるMEP Gデータ1のどの部分が、検索対象に対応するHMMモ

50

デルに対して尤度最大となるかを、M E P G データ 1 の中をスキャンして検索を行う。

【0100】この場合、効率的に M E P G データ 1 の中の最大尤度部分を求めるには、前記文献（ホ）に記載されている H M M スポットティングアルゴリズムを使用することが可能である。

【0101】以上の処理フローから明らかなように、H M M モデルによる認識は最尤推定により行われ、また、学習は、学習用データからの H M M モデルのパラメータの推定という形で実現される。

【0102】そして、シンボル系列全体から尤度計算が行われるため、カテゴリに特有のシンボル列パターンが現れていれば、時間軸方向の多少の移動、伸縮等に対して強いというメリットがある。

【0103】また、動画像の時系列パターンの各時点までの尤度を求め、これに対して閾値処理等を施すことで特定の時系列パターンの検索が可能となる。

【0104】次に、本発明の実施の形態に基づく実験結果例として、テニス動作画像を対象とした 2 つの人物動作確認実験結果について説明する。

【0105】〔実験 1〕本発明の実施の形態において、実験 1 に使用したテニス動作画像の写真の一例を図 5 に示す。

【0106】図 5 の上段に示すテニス動作画像から、図 5 の下段に示すように、背景差分により人物領域を抽出し、この人物領域が抽出された画像例をもとに作成した M E P G データを認識対象とし、D C T 計数を特徴量としたときの認識性能を評価した。

【0107】認識性能は、各ブロック（8×8 画素）当たりの D C T 係数を、低次成分から順に 1 列ずつ、即ち、1、3、6、10、15、21、28 個抽出して、それぞれ実験を行い、認識率を求めた。

【0108】なお、各ブロック当たりの D C T 係数が 1 のときは、D C 成分のみとなる。

【0109】また、画像サイズは、16×16 画素（マクロブロック単位で 1×1 ブロック）、32×32 画素（マクロブロック単位で 2×2 ブロック）の 2 種類とした。

【0110】また、量子化のためのコードブックのサイズは、各クラスサイズ 8、6 クラス合計で 48 とし、L B G アルゴリズムにより作成し、H M M モデルの状態数は 12、シンボル数は 48 である。

【0111】図 6 は、本発明の実施の形態の実験 1 で対象とするテニス動作画像を示す写真である。

【0112】図 6 に示すように、対象としたテニス動作は、バックハンドボレイ（back-volley）、バックハンドストローク（back-stroke）、フォアハンドボレイ（fore-volley）、フォアハンドストローク（fore-stroke）、スマッシュ（smash）、サービス（service）の

6 カテゴリである。

【0113】6 カテゴリの動作のそれぞれについて、10 試行の動作画像データを収集し、このうちの 5 試行を学習用データとして使用し、H M M モデルのパラメータの推定を行い、残りの 5 試行をテストデータとして、認識実験を行った。

【0114】この場合に、10 試行のうちから 5 試行を選択する選択方法を 10 通りに変えて実験を行った。

【0115】したがって、認識率は、 $5 \times 10 \times 6 = 300$ 回の認識実験のうち何回成功したかで評価される。

【0116】この認識実験結果を、表 1、表 2 に示す。

【0117】

【表 1】

表 1: 認識実験結果 (16x16)

DCT(per Block)	認識率 (%)
1	72.66
3	93.00
6	97.66
10	99.33
15	96.00
21	98.00
28	96.33

【0118】

【表 2】

表 2: 認識実験結果 (32x32)

DCT(per Block)	認識率 (%)
1	88.33
3	88.66
6	98.00
10	98.00
15	99.66
21	100.00
28	99.66

【0119】表 1、表 2 から理解できるように、特徴量として使用する D C T 係数を増やすことにより、認識率が大きく向上しており、比較的 low 周波成分の D C T 係数が人物動作の画像認識のための特徴量として有効であることがわかった。

【0120】また、対象画像が比較的小さい場合でも、D C T 係数を高周波成分まで使用することにより、98% 以上の認識率が得られ、画像が大きい場合と遜色のない認識率を実現できることがわかった。

【0121】〔実験 2〕本発明の実施の形態において、複数種の動作を含む一連の動画像データを対象として、動画像検索への適用実験を行った。

【0122】各動作カテゴリの学習済 H M M モデルによ

り、各時点までの尤度最大のHMMモデルを選ぶことにより、動作の検索が行えるかを検討した。

【0123】画面サイズは32×32画素を使用し、特徴量としてDCT係数は各ブロック当たり6とした。

【0124】図7は、本発明の実施の形態における、実験2の実験結果を示すグラフである。

【0125】図7は、各時点までの観測に基づいて、それぞれ6カテゴリのHMMモデルの対数尤度をプロットしたグラフである。

【0126】したがって、尤度は、動作の終了時に最大となることが期待される。

【0127】図7に示すグラフから、各対象動作のHMMモデルが順に最大尤度となっていることが確認でき、閾値処理により動作区間の切り出しが可能であることが理解できる。

【0128】これにより、連続動画データの中の特定の動作パターンの検索が可能である。

【0129】なお、前記本発明の実施の形態の説明においては、MEPG、MEPG2等の標準符号化方式により符号化されたMEPGデータを使用したが、これに限定されるものではなく、例えば、motion-JPEG等の標準符号化方式により符号化されたデータを使用できることはいうまでもない。

【0130】以上、本発明を発明の実施の形態に基づいて具体的に説明したが、本発明は、前記発明の実施の形態に限定されるものではなく、その要旨を逸脱しない範囲において種々変更し得ることはいうまでもない。

【0131】

【発明の効果】本願で開示される発明のうち、代表的なものによって得られる効果を簡単に説明すれば、下記の通りである。

【0132】(1) 本発明によれば、特徴量としてDCT係数、あるいは、DCT係数および動き補償ベクトルを使用するようにしたので、MEPG、MEPG2等の標準符号化方式で圧縮された少容量の動画データから、特定の動画パターンを直接認識・検索することが可能となる。

【0133】これにより、データ処理の処理時間を少な

くすることが可能となる。

【0134】(2) 本発明によれば、特徴ベクトル系列全体から尤度計算が行われるため、カテゴリに特有の特徴ベクトル列パターンが現れていれば、時間軸方向の多少の移動、伸縮等があっても、特定の動画パターンを精度良く認識・検索することが可能となる。

【0135】(3) 本発明によれば、特徴量として使用するDCT係数を高周波成分まで使用することにより、認識率を大幅に向上させることができ、また、対象画像が比較的小さい場合であっても、特徴量として使用するDCT係数を高周波成分まで使用することにより、認識率を向上させることが可能である。

【0136】(4) 本発明によれば、銀行や商店における不審行動監視、スポーツなどの動画から所望の動作部分の切り出しなどに広く適用できる。

【図面の簡単な説明】

【図1】本発明の一発明の実施の形態である動画認識方法および動画認識検索方法が適用される動画認識検索装置の概略構成を示す機能ブロック図である。

【図2】MEPGデータ1およびMEPGデータ1のDCT係数の概略構成を示す図である。

【図3】本発明の実施の形態における、DCT係数の抽出方法を説明するための図である。

【図4】HMMモデル（隠れマルコフ）の概念を示す概念図である。

【図5】本発明の実施の形態において、実験1に使用したテニス動作画像の例を示すディスプレイ上に表示した中間調画像である。

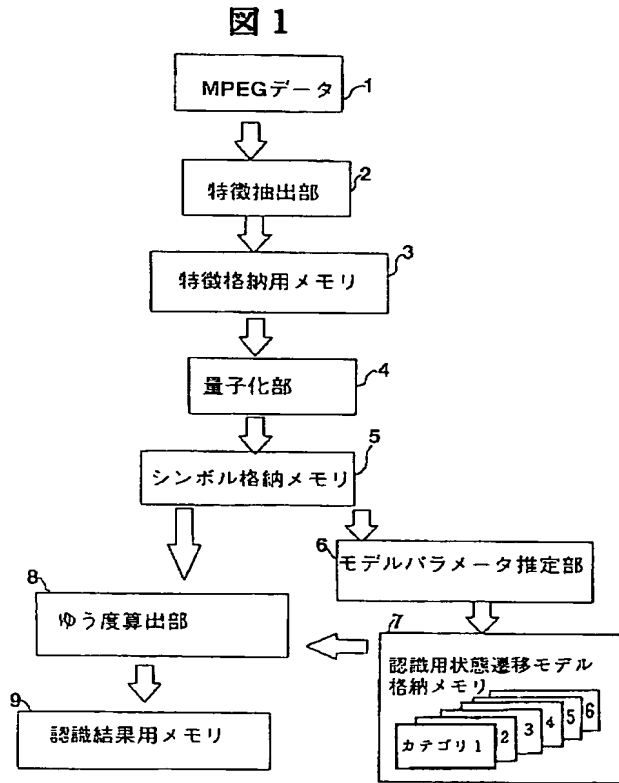
【図6】本発明の実施の形態の実験1で対象とするテニス動作画像を示すディスプレイ上に表示した中間調画像である。

【図7】本発明の実施の形態の実験2の実験結果を示すグラフである。

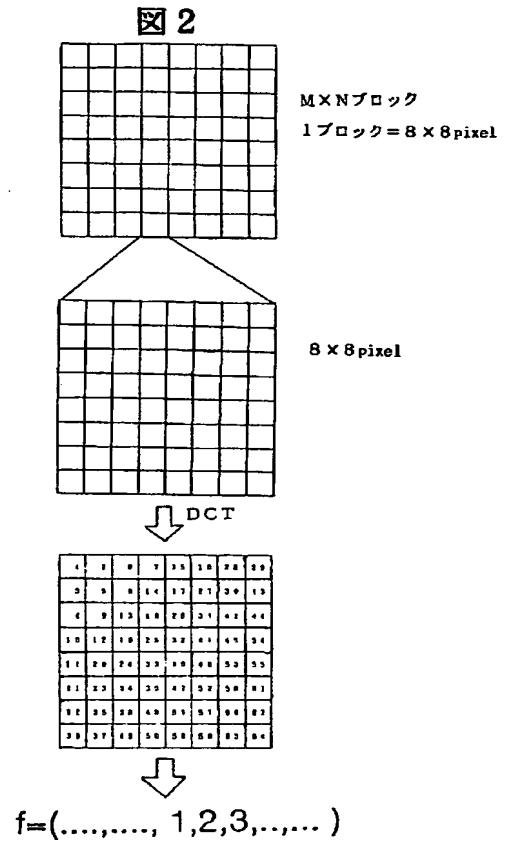
【符号の説明】

2…特徴抽出部、3…特徴格納用メモリ、4…量子化部、5…シンボル列格納用メモリ、6…モデルパラメータ推定部、7…認識用状態遷移モデル格納メモリ、8…尤度算出部、9…認識結果用メモリ。

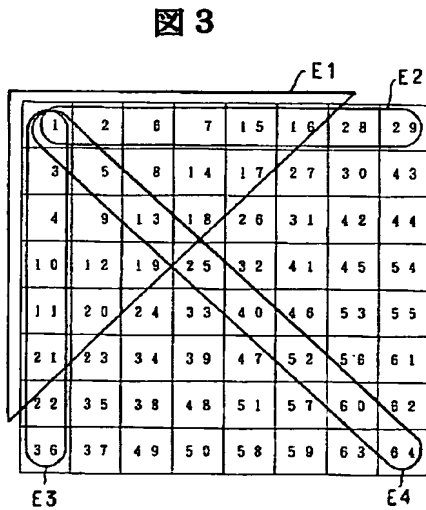
【図1】



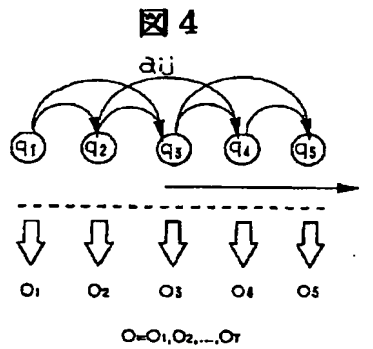
【図2】



【図3】

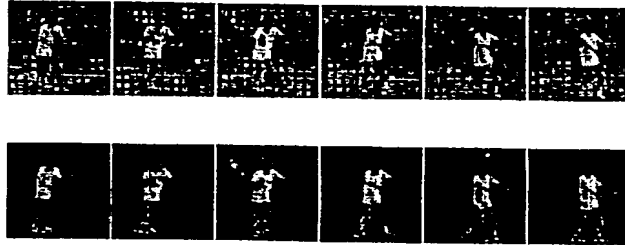


【図4】



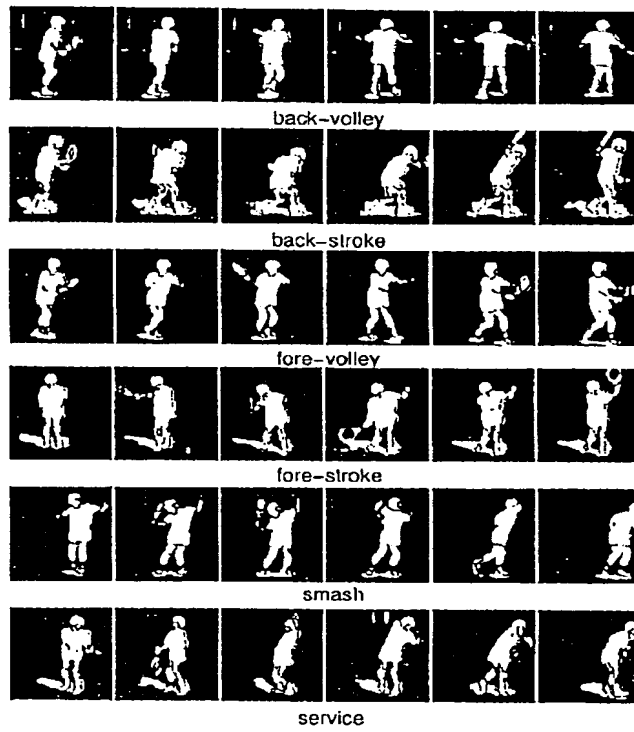
【図 5】

図 5



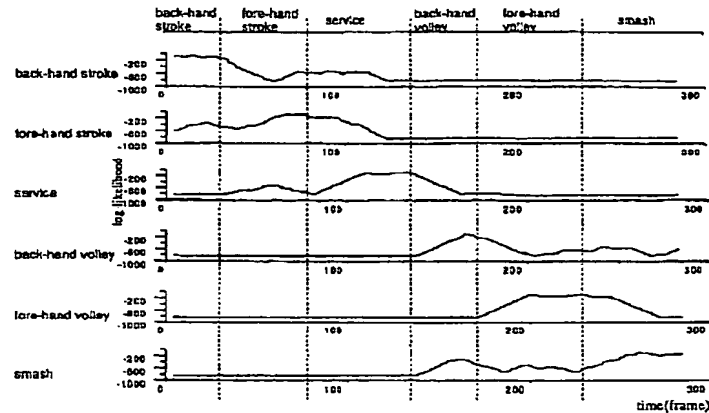
【図 6】

図 6



【図7】

図7



【手続補正書】

【提出日】平成8年6月27日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正内容】

【特許請求の範囲】

【請求項1】 一連の動画像の動画像パターンを認識する動画像認識方法において、一連の動画像を表示する各画面の画像データを $M \times N$ のブロックに区切り、各ブロックのDCT係数を抽出するステップと、前記DCT係数の少なくとも1つを各画面の特徴ベクトルとして抽出するステップと、特定の動画像パターンを表示する各画面の特徴ベクトルで構成される時系列の特徴ベクトル列により、確率的な状態遷移モデルを、認識キーとなる複数の特定の動画像パターン毎に学習するステップと、認識対象である一連の動画像を表示する各画面の画像データから抽出された特徴ベクトルで構成される時系列の特徴ベクトル列の、前記学習により得られた複数の状態遷移モデルに対する尤度が最大となる状態遷移モデルの動画像パターンを認識結果として出力するステップとを具備することを特徴とする動画像認識方法。

【請求項2】 前記認識対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の一部を使用することを特徴とする請求項1に記載された動画像認識方法。

【請求項3】 前記各画面の特徴ベクトルとして、DCT

T係数とともに動きベクトルを使用することを特徴とする請求項1に記載された動画像認識方法。

【請求項4】 前記認識対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の一部、および、動き補償ベクトルを使用することを特徴とする請求項3に記載された動画像認識方法。

【請求項5】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の内、3ないし21個の低周波成分のDCT係数を特徴ベクトルして使用することを特徴とする請求項2または請求項4に記載された動画像認識方法。

【請求項6】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の内、水平方向の1番目のライン上のDCT係数を特徴ベクトルして使用することを特徴とする請求項2または請求項4に記載された動画像認識方法。

【請求項7】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の内、垂直方向の1番目のライン上のDCT係数を特徴ベクトルして使用することを特徴とする請求項2または請求項4に記載された動画像認識方法。

【請求項8】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれるDCT係数の内、直流成分を含む対角線上のDCT係数を特徴ベクトルして使用することを特徴とする請求項2または請求項4に記載された動画像認識方法。

【請求項 9】 一連の動画像の中から、特定の動画像パターンを含む時間領域を抽出する動画像認識検索方法において、一連の動画像を表示する各画面の画像データを $M \times N$ のブロックに区切り、各ブロックの DCT 係数を抽出するステップと、前記 DCT 係数の少なくとも 1 つを各画面の特徴ベクトルとして抽出するステップと、検索キーとなる特定の動画像パターンを表示する各画面の特徴ベクトルで構成される時系列の特徴ベクトル列により、確率的な状態遷移モデルを学習するステップと、検索対象である一連の動画像を表示する各画面の画像データから抽出された特徴ベクトルで構成される時系列の特徴ベクトル列の中で、前記学習により得られた状態遷移モデルに対する尤度が高い時間領域を検索結果として出力するステップとを具備することを特徴とする動画像認識検索方法。

【請求項 10】 前記検索対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部を使用することを特徴とする請求項 9 に記載された動画像認識検索方法。

【請求項 11】 前記各画面の特徴ベクトルとして、DCT 係数とともに動きベクトルを使用することを特徴とする請求項 9 に記載された動画像認識検索方法。

【請求項 12】 前記検索対象である一連の動画像を表示する各画面の画像データが、標準符号化方式により圧

縮されており、各画面の特徴ベクトルとして、標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の一部、および、動き補償ベクトルを使用することを特徴とする請求項 11 に記載された動画像認識検索方法。

【請求項 13】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、3 ないし 21 個の低周波成分の DCT 係数を特徴ベクトルとして使用することを特徴とする請求項 10 または請求項 12 に記載された動画像認識検索方法。

【請求項 14】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、水平方向の 1 番目のライン上の DCT 係数を特徴ベクトルとして使用することを特徴とする請求項 10 または請求項 12 に記載された動画像認識検索方法。

【請求項 15】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、垂直方向の 1 番目のライン上の DCT 係数を特徴ベクトルとして使用することを特徴とする請求項 10 または請求項 12 に記載された動画像認識検索方法。

【請求項 16】 前記標準符号化方式により圧縮された各画面の画像データ中に含まれる DCT 係数の内、直流成分を含む対角線上の DCT 係数を特徴ベクトルとして使用することを特徴とする請求項 10 または請求項 12 に記載された動画像認識検索方法。